Educational Equity in New York

Project Technical Report

Paul Polsinelli, Stanley Perez, Tavianne Kemp, Yiqing Guo

Introduction:

Backstory:

In this project, we aim to investigate educational equity in school districts in New York State. Using graduation rate as a metric for educational outcomes, we aim to analyze educational outcomes compared to the district's student body information and financial status. Student body information includes total enrollment, percentage of minority students, percentage of graduation and dropout rate. Financial status includes federal funding status, funding per student, and median household income for that district.

We chose the topic of educational equity because educational outcomes could play a major role over the course of a student's life. Trying to understand what factors can affect a student's educational outcome is important because it can help educational leaders understand which areas need to be improved upon.

Exploratory questions:

1. Which districts have the most funding per student and what differentiates them from the other districts? Is there a difference in graduation rate for these districts compared to the ones that don't receive as much funding?
2. What is the percentage of districts that get federal funding vs ones that do not? How do the districts that received federal funding compare to the ones that didn't?
3. What kind of correlation is there between the graduation rate and the other factors we've acquired?
4. Is there a discrepancy between the graduation rate and the percentage of minority students in a district?
5. Can we predict the graduation rate for a district based on the data and which factor has the most effect on the graduation rate for a school district?

Research:

We began research by looking at worldwide educational data. Because of the sheer number of variables that can affect education in different countries worldwide, we decided to narrow the scope of the research to just New York State, a region with

educational standards and funding distribution supervised by one government entity, the New York State Education Department (NYSED).

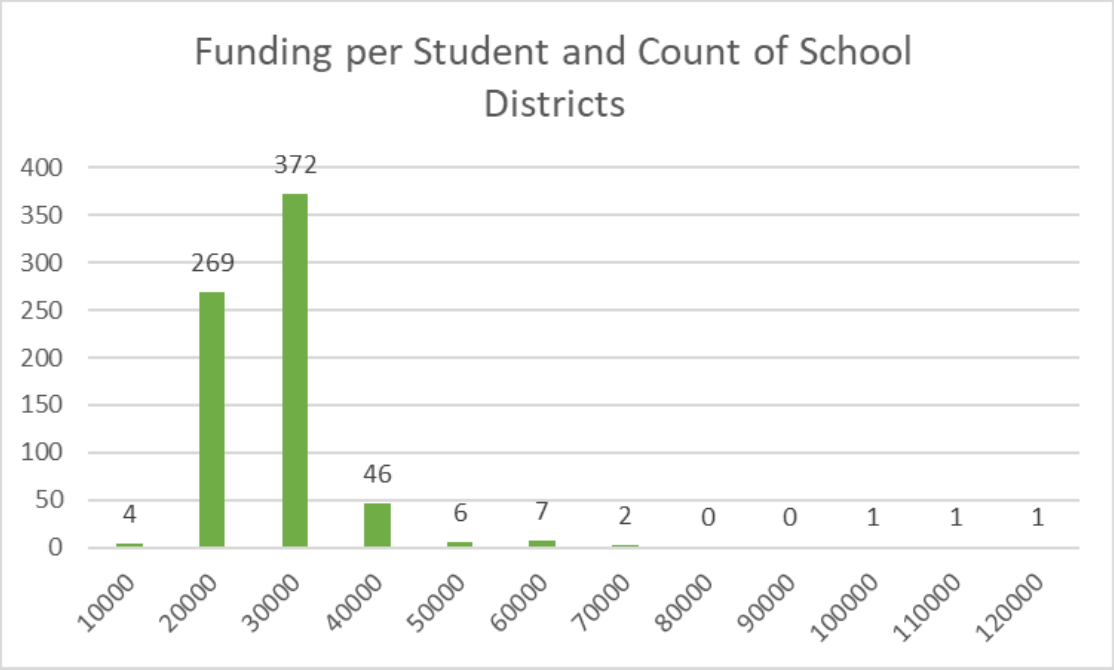
The NYSED releases data on graduation rate, dropout rate, and other outcome data for each school per district on the data.nysed.gov website. The website also includes student body demographics and the amount of money spent on each student per district. The US census Bureau provides median income by zip code, which, when combined with a table to convert zip codes to school district, gave us a dataset for the median household income per school district. We then found data from the NYSED that provided us with budget information, especially whether federal funding was provided to each school district.

One data cleaning issue that we encountered was that NYC was not separated out by district in our federal funding dataset. To keep NYC on a district level we found separate sources that provided the total funding per NYC school district.

Exploratory Data Analysis:

To assess whether the datasets are usable for the purposes of this project, we conducted exploratory data analysis on the datasets. After leaving only relevant columns, we assessed data columns for frequencies (total amount of school districts) in a histogram.

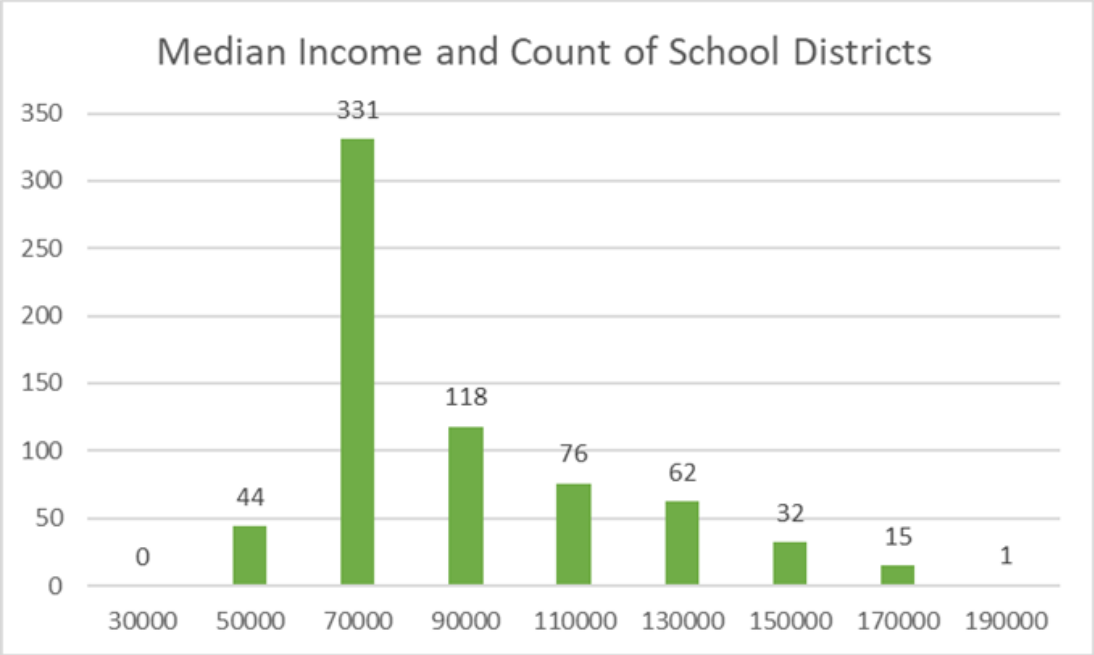
From the web scraped district data, we found information on total funding per student for each school district. In the following histogram (Figure 1), the funding per student is organized into bins of $10000 ranges per bin, where the value on the x axis represents the maximum dollar amount in that bin. Most school districts spend between $10000 and $30000 per student.



[Figure 1] - School districts within a range of student funding

- Average dropout rate: 4.77%
- Average graduation rate: 88.5%
- Average percentage of minority students: 26.9%

From the median income census dataset, which was converted from zip code data into district data using SQL joins, we created the following histogram (Figure 2). This bar chart represents the distribution of school districts by household median income bins of $20000 ranges. Each number on the x axis represents the maximum value of the $20000 bin. A little less than half of all school districts have communities with a household median income of $50000 to $70000.



[Figure 2] - School districts within a range of median income

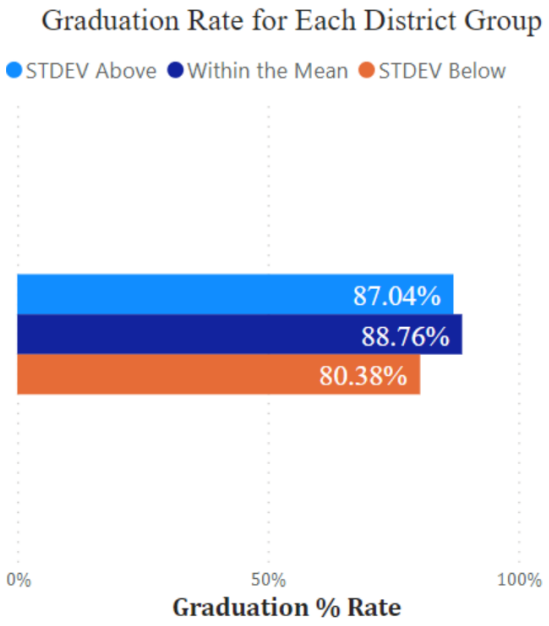
- Average median income for all districts: $77,873

Discussion:

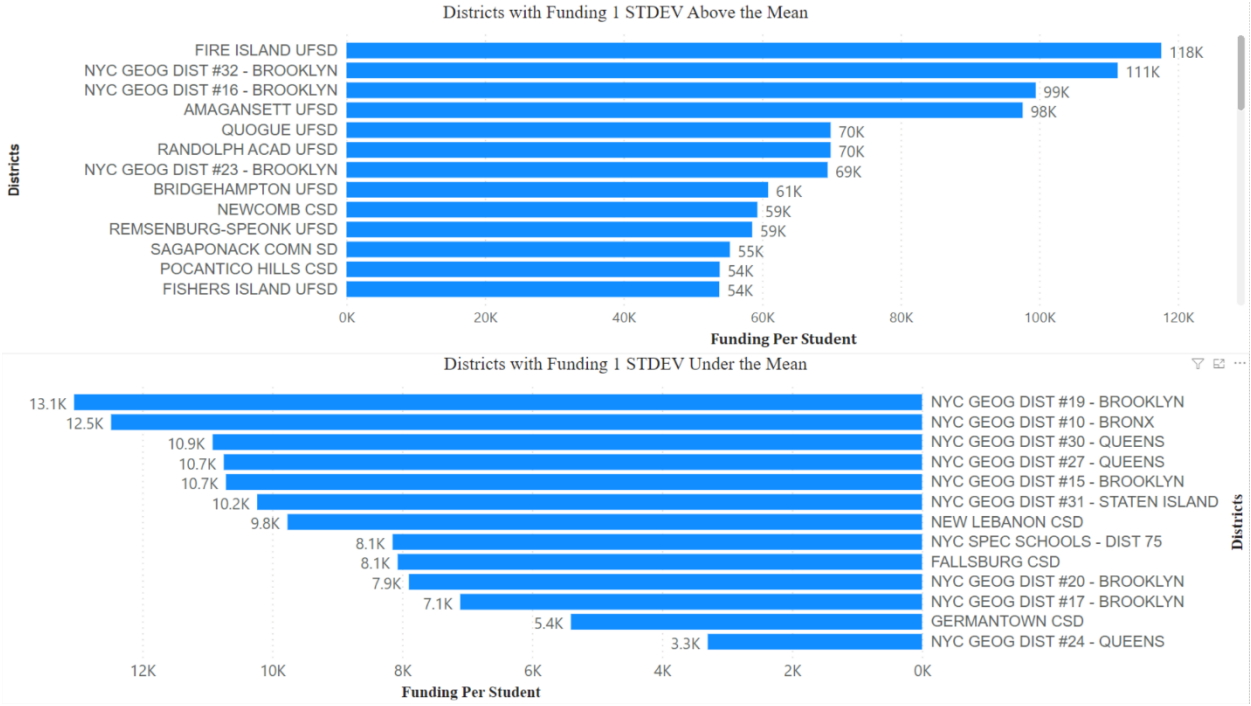
Wraps up research and ties back to exploratory questions

1. Which districts have the most funding per student and what differentiates them from the other districts? Is there a difference in graduation rate for these districts compared to the ones that don't receive as much funding?

Figure 3 outlines the average graduation rate for districts +/- 1 standard deviation above the mean in terms of funding per student. The average funding per student in a NY State school district is about $23,251.60. The average graduation rate for all schools with funding over $32,873 is 87.04% and all the schools that have funding less than $13,630.18 is 80.38%. All the schools that are within 1 standard deviation average 88.76% showing that the districts with average funding perform better than the districts that receive the most funding.



[Figure 3] - Bar chart outlining average graduation rate for each district group.



[Figure 4] - Top graph outlines the districts with the highest funding. Bottom graph shows the districts with the least funding per student.

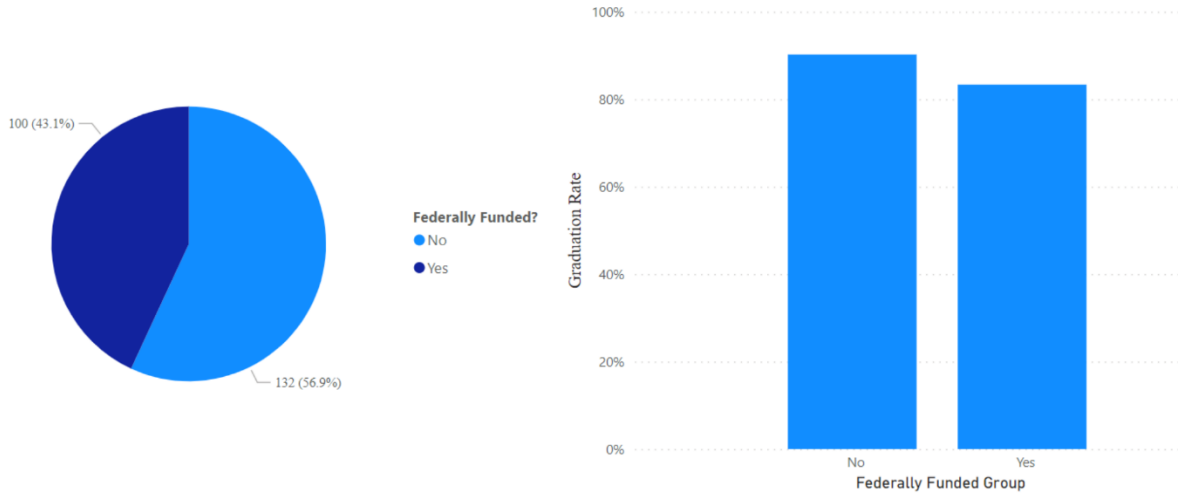
Taking a closer look at the districts with the most funding (Figure 4 Top Graph), the school district with the most funding is Fire Island UFSD. This school district spends $117,684 per student and has a graduation rate of 96%. Some schools in the high funding per student category are like Fire Island UFSD, in that they have a high funding

per student and a high graduation rate. However, the NYC Geographic Districts do not follow this trend. While NYC Geographic District #32 spends $111,293 per student, it has a graduation rate of 75%.

The NYC Board of Education is most likely distributing funding based on student needs. NYC School Districts show up in both the high funding and low funding category bar charts. However, the performance of NYC School Districts in the high funding bar chart is noticeably lower than the performance of NYC School Districts in the low funding bar chart. A particularly extreme example of this is NYC Geographical District #23, which has $69,436 in funding per student, with a graduation rate of 52%. On the other hand, the NYC Geographical Districts in the low funding category bar chart have noticeably lower funds per students, ranging from $13,050 to $3,298, and graduation rates closely clustered around the mean graduation rate of 80.38%.

In terms of balancing the graduation rates, it would make sense to distribute the funds more equally across districts given that the districts with the most funding have a slightly lower average graduation rate than districts with average funding. This can possibly indicate that after a certain point in funding, it doesn't have a significant effect on the graduation rate. On the other side, the districts with the least funding show a significant decrease in graduation rate (7%) and after running a two-sample t-test *($p < .01$)* on the subset groups, the test concludes that this difference in percentage points didn't happen by chance and there's a statistical significance that warrants further research. Distributing the funds from the high end to the low end could have a positive effect in trying to balance out the graduation rates across all districts.

2. What is the percentage of districts that get federal funding vs ones that do not? How do the districts that received federal funding compare to the ones that didn't?



[Figure 5] - Percentage of districts that are federally funded vs districts that aren't (left). Graduation rate for each group (right).

Breakdown of Stats by Federal Funding Status

Federal Funding Status	Avg Minority Student Percent	Avg Enrollment	Avg Graduation Rate	Avg Funding per Student	Avg Household Income
	19.67%	1,431.52	89.09%	24,046.32	74,824.98
No	35.71%	4,104.39	90.23%	21,962.98	95,075.41
Yes	50.45%	11,889.16	83.36%	21,122.05	68,754.33
Total	**26.94%**	**3,390.32**	**88.50%**	**23,251.60**	**77,873.66**

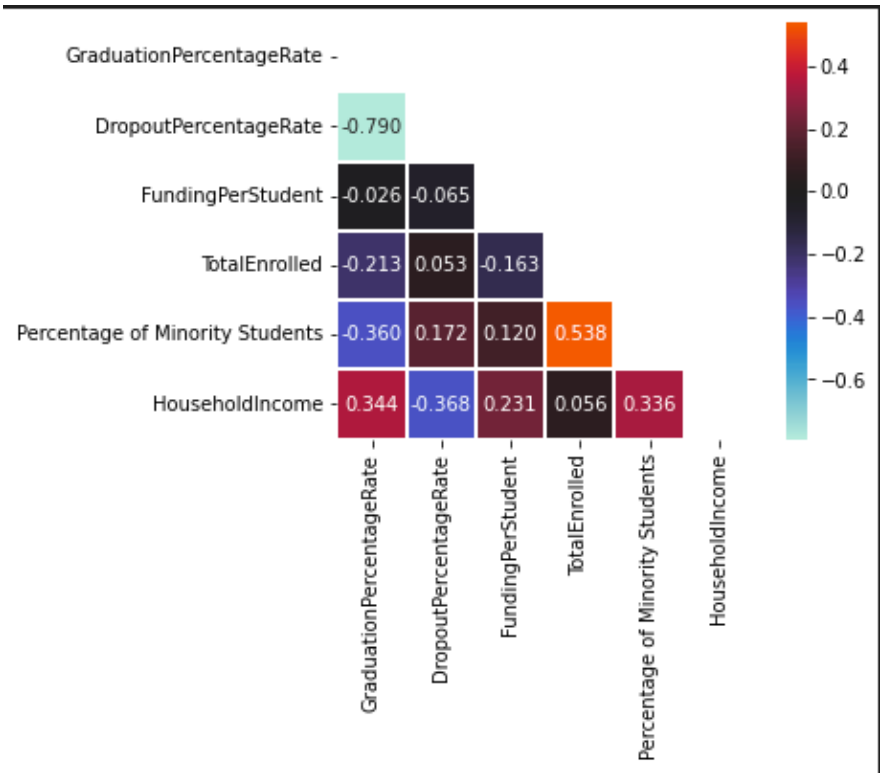
[Figure 6] - Pivot table displaying difference between federally funded districts vs non federally funded.

The dataset of federally funded schools contains only 223 school districts in New York State. 487 school districts do not contain information on whether they are federally funded or not. Of the 223 school districts included, a little under half of them receive federal funding, while the rest do not receive federal funding.

Looking at the bar chart, the average graduation rate of federally funded school districts is lower than that of non-federally funded school districts by about 7%. This was further supported with a two-sample t-test *($p < .01$)* showing that this statistical difference didn't happen by random chance. Furthermore, looking at the pivot chart of average percent of minority students, average funding per student, average household income by federal funding status, the federally funded school districts have more minority students (25%), lower funding per student (by $2700 on average), and lower household income (by $26500 on average) as well. This shows that federal funding is most likely to be distributed according to the needs of the school district. Federal funding goes to support schools with less graduating students, where the average household income is a lot lower. Looking deeper into the data we see that districts that are non-federally funded enrolled about 4,817 students for the year 2019-2020. This pales in comparison to the 11,889 students enrolled in federally funded districts. This stark difference could also play a factor in the graduation rate due to the need for teachers per student.

Further research would be warranted but districts that are federally funded enrolling more students would need to hire more teachers. Sometimes more students are allocated to each teacher instead, which could make it difficult for a teacher to give the time and attention to each student to ensure they understand the material. It could lead to a trickledown effect in that more students per class leads to less time and attention to each student leading to students not fully understanding the material which could lead to a drop in grades that lead to less students graduating.

1. What kind of correlation is there between the graduation rate and the other factors we've acquired?

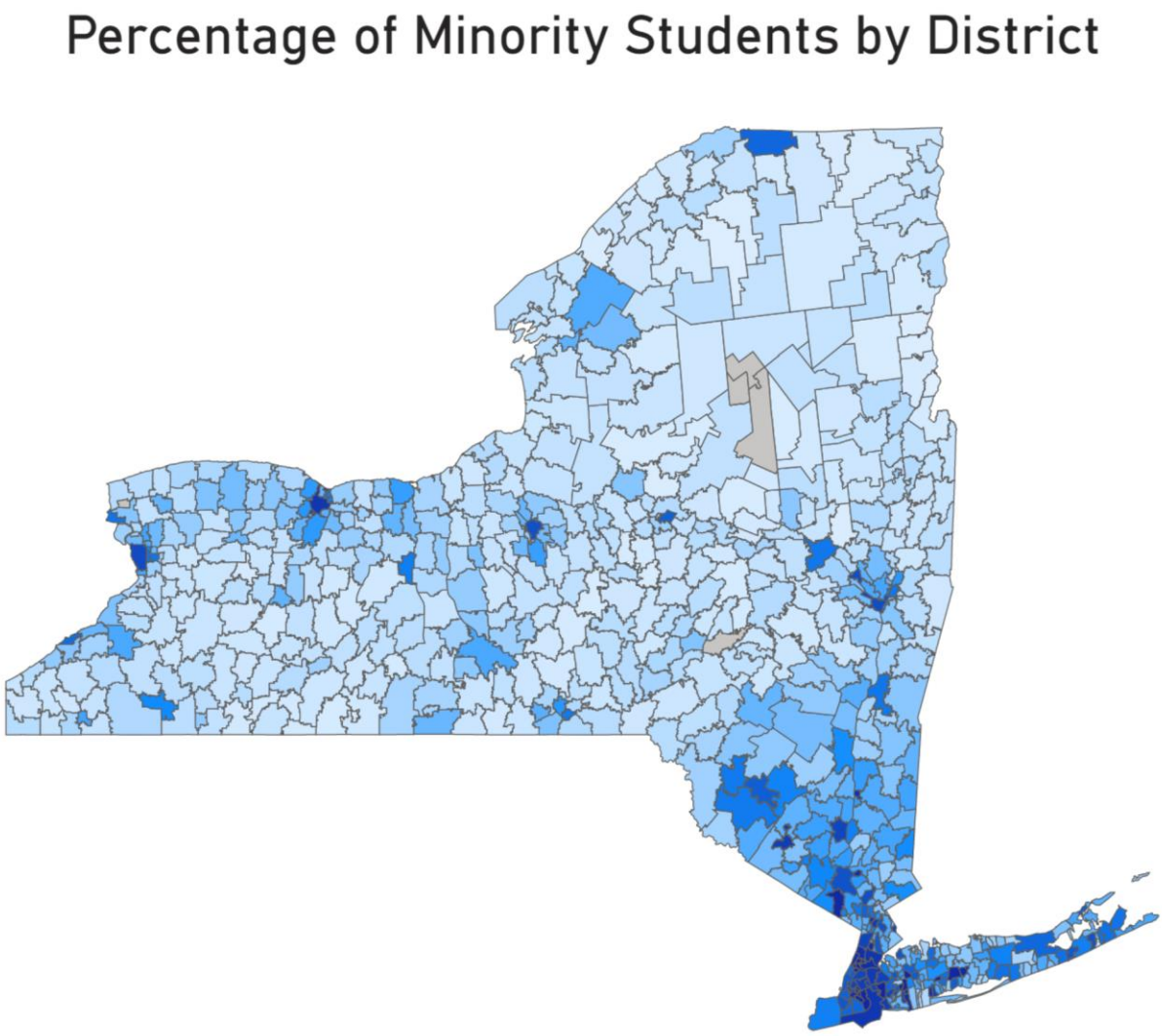


[Figure 6] - Correlation matrix between the different factors acquired for each district.

According to this correlation matrix, the items with highest correlation are total amount of students and graduation rate, total enrollment and funding per student, percent of minority students and graduation rate, percent of minority students and total enrollment, graduation rate and household income, funding per student and household income, and percent of minority students and household income.

Total enrollment for a school district can be thought of as an indicator for how sparsely populated a region is, or if it is urban, suburban, or rural. From this map of school districts in New York State and percentage of minority students, it can be seen that the darkest blue areas, with the highest minority students, center around cities and urban centers, going from around 80% in cities such as Buffalo City, Rochester City,

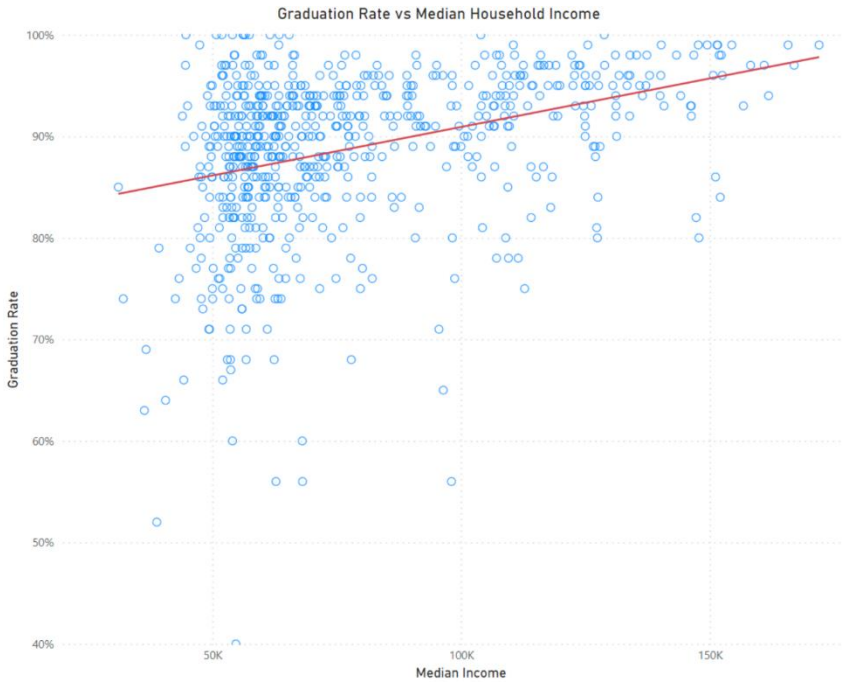
Syracuse City, with the highest concentration around New York City, to close to 0% in other parts of the state. Overall, the correlation between total enrollment and percent of minority students shows that there is a higher minority population in cities and other densely populated areas as compared to more rural areas.



[Figure 7] - Heat map showing the areas with the most minority students.

There is a relatively high negative correlation between total enrollment and graduation rate. When taken into account with the similar negative correlation between total enrollment and funding per student, it seems that as the total number of students in a district increases, the less resources and support can be provided to each student.

Another interesting correlation is everything involving household income. A higher household income correlates positively with graduation rate, minority student percentage, and funding per student. It also correlates negatively with dropout rate.



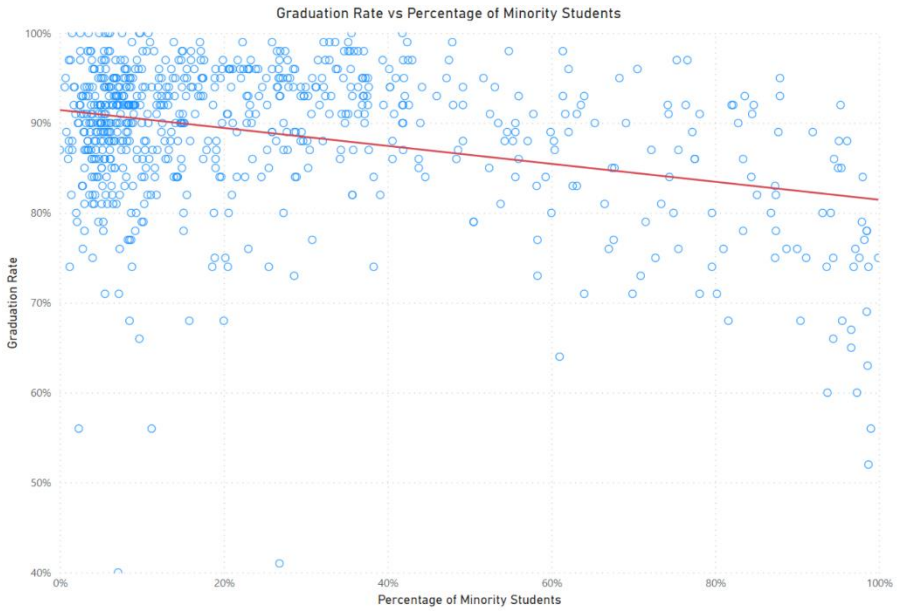
[Figure 8] - Graduation rate vs Median Household income for each district

This scatter plot of household income vs graduation rate shows this positive trend. In districts with richer families, the graduation rate increases. Going by this chart alone, however, is not enough data to show that the increase in graduation rate is an isolated effect of a rise in household income.

The correlation with household income and funding per student can be attributed to how schools (especially not in the cities) are majority funded by local revenue, which consists of taxes, mostly property taxes (NCES). As an area is richer or the land is evaluated at a higher price due to desirability, the amount of money funding each school increases, which increases the funding per student.

The correlation with household income and graduation rate could be because of a few reasons. Students in poorer families may start working sooner or have more need to support themselves and their families. Time spent in schooling is a lost opportunity cost when compared to immediately earning money by working a job. Students living in poorer communities could be exposed to more dangerous conditions than their richer counterparts, and will be less able to pay for resources or outside help like tutoring. Living in impoverished neighborhoods leads to a higher risk of mental illness and adverse medical outcomes (Georgia Tech). These factor can all affect student performance in schools.

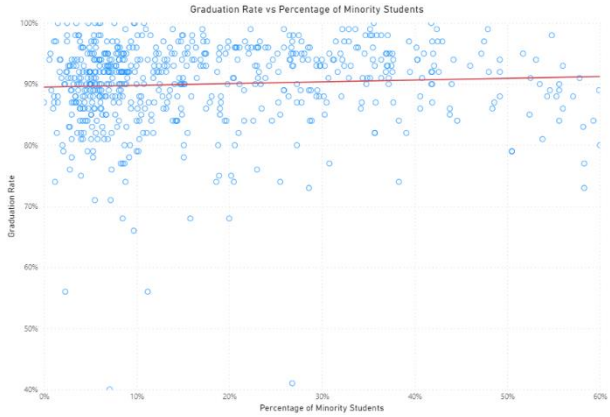
2. Is there a discrepancy between the graduation rate and the percentage of minority students in a district?



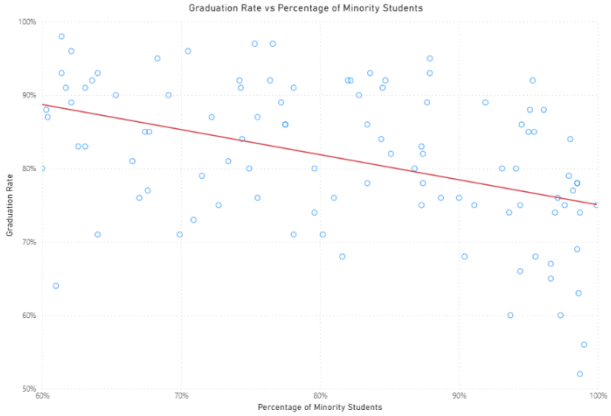
[Figure 9] - Graduation Rate vs Percent of Minority Students in a district

The scatter plot of percentage of minority students vs graduation rate shows a negative trend, showing that in districts with more minority students, the graduation rate decreases. There is not enough data in this plot to say that this is a causation, seeing as higher percentages of minority students also correlates with lower household median income, and lower funding per student, as shown previously.

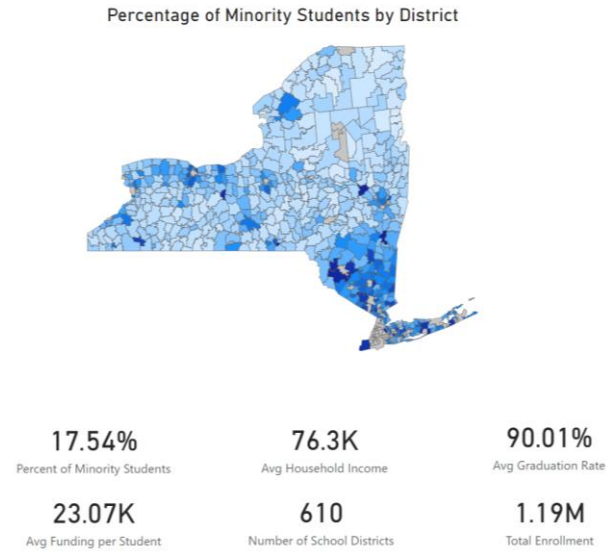
When this graph is sliced by percentage of minority students, we can see that a major shift takes place in the trendline after minority percentage increases over 60%. While the trendline is generally stable and even shows a slight positive inclination from 0% to 60%, the trendline begins to shift downwards from 60% to 100%.



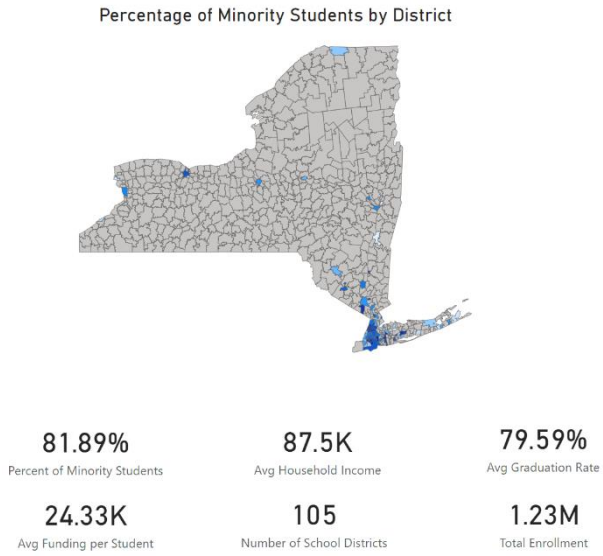
[Figure 10] - Graduation Rate vs Districts with <60% minority students



[Figure 11] - Graduation Rate vs Districts with >60% minority students



[Figure 12] - Heat map for districts with <60% minority students



[Figure 13] - Heat map for districts with >60% minority students

The map visualization shows the stark differences between school districts with <60% minority percentage and those with >60% minority percentage. As discussed previously, large urban centers like New York City, Buffalo City, and Syracuse City, among others, have the highest percentage of minority students in the state.

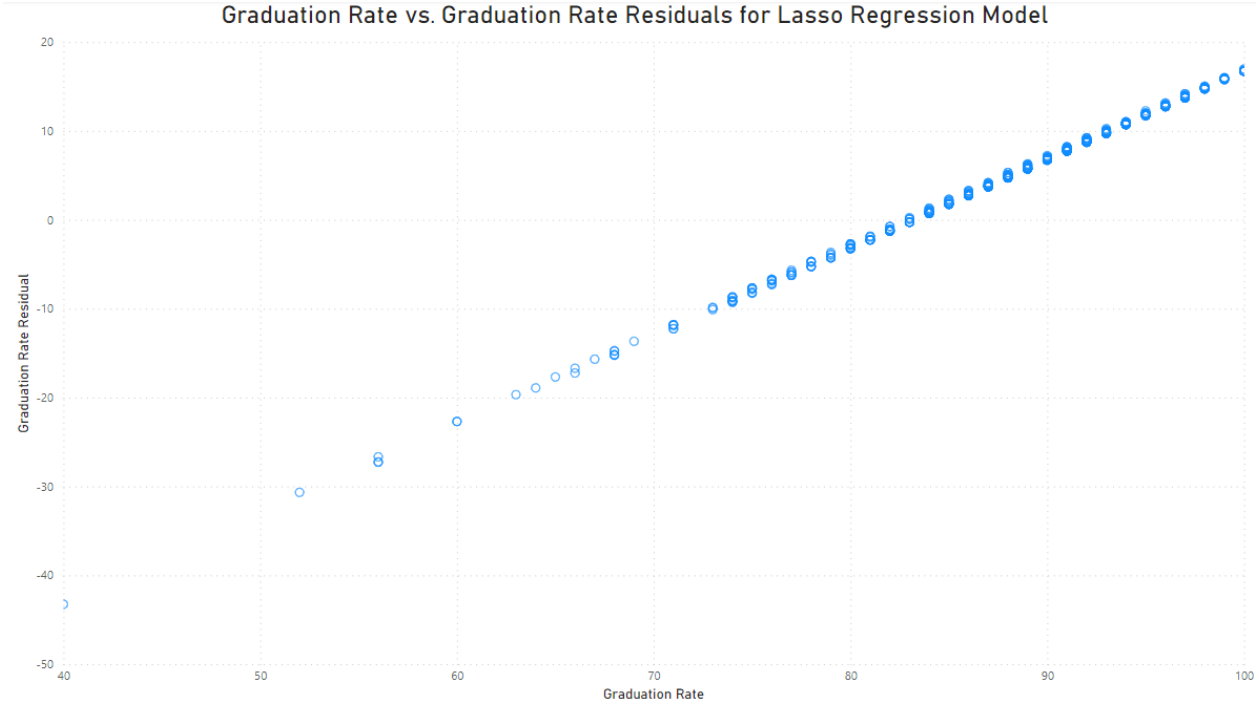
The statistics for both subsets of the dataset provide us with interesting observations as well. One observation is how the breakout of minority percentage brackets bucks the trend of more household income leading to higher graduation rates. When comparing the two groups, the <60% minority percentage map has a lower average household income by $10,000, and a higher graduation rate by 10%. This discrepancy, according to our previous observations, could be attributed to the higher cost of living and especially rent present in urban areas such as New York City compared to the rest of the state. Additionally, our household income dataset comes from the US Census data for all households, and not just households with children currently enrolled in a public school district. This could be skewing our data with the addition of non-student households. Further study should be conducted on the effects of household income adjusted for the region's cost of living, as well as focusing the dataset on households that support students enrolled in each specific school district.

Another observation is how the <60% minority percentage map has six times the number of school districts as the >60% minority percentage map, and yet school districts in the >60% minority percentage bracket enroll more students and receive more funding. This highlights the sheer scale of urban populations compared to surrounding rural regions, as well as the increase in difficulty for school districts in urban areas to keep up with six times as many students as other districts in the state. Urban school

districts could also be dealing with more students per classroom than their rural peers, leading to a decrease in student and teacher interactions and less understanding of the material. Additionally, continuing with the observation that cost of living in urban areas is higher, the same amount of funding will go a lot less far in urban school districts than it does in rural school districts. This could mean less resources per student, or less teachers are able to be hired at New York City salaries compared to New York State salaries. Further research should be done on the specific ways that urban and rural school districts differ, as well as classroom sizes and teacher salaries in urban and rural areas.

3. Can we predict the graduation rate for a district based on the data and which factor has the most effect on the graduation rate for a school district?

Our predictions showed correlation to actual graduation rate data of school districts based on the socio-economic predicting factors of household income, funding per student, and percentage of minority students. When we plotted the residuals, we found that the predictions were within approximately 5.7% of the actual graduation rate on average. The R^2 value of our model was only 19.4% with optimal hyper-tuning. What we can glean from this is the complexity of the problem that the factors we chose to aid in predicting graduation outcomes were insufficient to address the problem of how to increase graduation rates.



[Figure 14] - Graduation rate residuals (y-axis) vs Graduation rate (x-axis)

Using the linear coefficients to measure the change of the independent variables, we were able to see the variable which had the most effect on graduation rates. The

percentage of minority students in the district had the highest effect on graduation rates, followed by household income and funding per student.

- Percentage of Minority Students: -16.7%
- Household Income: 0.016%
- Funding Per Student: -0.011%

For our model, we used Lasso regression with a grid search to find the optimal settings for our parameters. Since our dataset was relatively small with four independent variables and six hundred and seventy-seven rows, the grid search was the most thorough and time efficient enough compared with a random search for optimization.

Conclusion:

In conclusion, we explored the relationships between different variables that led to both funding levels and graduation rates in school districts in New York State. One observation we made was that many of the variables are not independent of one another. For example, total enrollment is often a metric of how urban or rural a school district was, and the map visualization for percent of minority students in New York State shows that minority student percentage is often highest in urban centers, where total enrollment is highest. Another connection between variables is that a higher household income generally leads to more funding per student. This is because property taxes in New York State fund schools, so more local revenue in taxes leads to more money for the schools.

The graduation rate generally positively correlates with household income and funding per student because more funding means more resources available for all students. Schools that receive federal funding (especially in NYC Geographical Districts) tend to perform worse in terms of graduation rate, and their students tend to have a higher minority percentage and lower household income than schools without federal funding. This leads to the conclusion that federal funding is distributed on a need's basis, to offset the disadvantages that school districts in poorer community's face.

The Lasso ML Regression model performed well in its prediction based on how it fits the independent variables. Its low R^2 value was revealing as to the overall impact and relevancy of the variables. The problem is very complicated and with the proper type and amount of data the model would conceivably increase its R^2 value and be an effective tool in predicting graduation rates and informing policy makers as to where they can make socio-economic improvements to improve educational outcomes.

Recommendations made

Figure 4 shows that districts that receive federal funding on the higher end have a graduation rate in the same vicinity as the districts that receive funding around the mean (87.04% vs 88.76%). The districts on the lower end of funding have an average graduation rate of 80.38%. Running a two-sample t-test analysis showed that this was not by chance and it's statistically significant. A possible recourse could be to distribute the funds more evenly since from the data we can see that even if districts are receiving more than the average in funding per student, it is not having a significant effect on the graduation rate. Allocating extra funding to the districts on the lower end could help balance the scales and even out the graduation rates.

Figure 13 shows that 105 districts enrolled 1.23 million students which is more students than the other 610 districts who enrolled 1.19 million students. The figure also shows that these students were relatively in the same urban areas. The 10% difference in graduation rate could be attributed to the idea that more students in one area would require more teachers to be hired. New York City for example has had issues with class sizes being too large and too many students allocated to one teacher. This could have a trickle-down effect in that more students per teacher could mean less time the teacher has with each student to make sure the material is understood. This could lead to the student not understanding the material which could lead to poor grades that could eventually lead to lower graduation rates. Further study is warranted.

We recommend more research on specific factors that affect the trends we observed between variables and graduation rate. The connections between variables like high minority populations in urban school districts should be further explored. We do not know the specific reasons that lead to high minority populations experiencing worse outcomes than low minority populations on average. Because there is also the additional factor of minority populations concentrated in urban areas, research should be done to establish the independent effects of urban areas on education. Is there a difference between how schools in urban vs rural areas operate, and what different challenges do students face in urban vs rural communities?

Additional research based on the questions raised by our observations in this project should be done on the effects of household income adjusted for the region's cost of living, and data should be collected on the household incomes of households that specifically support students enrolled in each specific school district. We also recommend research into how urban and rural school districts differ in management, teacher salary levels, and classroom sizes.

Lastly, we cannot discount the effects of COVID-19 on graduation rates since our dataset references the 2019-2020 school year. This school year was interrupted by COVID-19 quarantine guidelines in March of 2020, which affected schooling for

everyone by forcing adoption of online instructional methods. This sudden change could have affected some students more than others, and interrupted graduation plans for many school districts and students. Further research should be done to isolate the effects of COVID-19 from the data.

Sources

Sielke, C. by C. C. (2001, March 19). *Public School Finance Programs of the U.S. and Canada: 1998-99*. National Center for Education Statistics (NCES) Home Page, a part of the U.S. Department of Education. Retrieved August 17, 2022, from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001309

NYC Class Size 2020. New York City Department of Education. (2020, February). Retrieved August 17, 2022, from https://infohub.nyced.org/docs/default-source/default-document-library/2019-20-february-class-size-report---webdeck---accessible.pdf

Create Shape Map Visualizations in Power BI Desktop (preview). Use Shape maps in Power BI Desktop (Preview) - Power BI | Microsoft Docs. (2022). Retrieved from https://docs.microsoft.com/en-us/power-bi/visuals/desktop-shape-map

Department of Health and Mental Hygiene (DOHMH). (2020, May 13). *Modified ZIP code tabulation areas (MODZCTA)*. Modified Zip Code Tabulation Areas (MODZCTA) | NYC Open Data. From https://data.cityofnewyork.us/Health/Modified-Zip-Code-Tabulation-Areas-MODZCTA-/pri4-ifjk

NCES. (2022). *School District Geographic Relationship Files*. Geographic. Retrieved from https://nces.ed.gov/programs/edge/Geographic/RelationshipFiles "\GRF21\grf21_lea_zcta5ce20.xlsx"

New York State. (2019). *2019 District Data – School Funding Transparency*. https://openbudget.ny.gov/sft/sft-districts-19.html

New York State. (2022). *GIS.NY.GOV*. NYS GIS Clearinghouse - NYS Education Department - NYS Schools and School District Boundaries. From https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1326

NYC Open Data. (2022). *School districts*. NYC Open Data. From https://data.cityofnewyork.us/Education/School-Districts/r8nu-ymqj

NYC OpenData. (2021). *FY2020 Local Law 16 Final Report*. https://data.cityofnewyork.us/Education/FY2020-Local-Law-16-Final-Report/cvqn-xqrr/data

NYSED. (2020). *Districts: NYSED Data Site*. data.nysed.gov. From https://data.nysed.gov/lists.php?type=district

Problem Using Custom Shape Map. Microsoft Power BI Community. (2021, September 13). Retrieved from https://community.powerbi.com/t5/Desktop/Problem-using-custom-shape-map/m-p/55795

U.S. Census Bureau. (2020). *Income in the Past 12 Months (In 2020 Inflation-Adjusted Dollars)*. https://data.census.gov/cedsci/table?q=median+income&g=860XX00US11701